Abstract

We survey different optimization problems under uncertainty which arise in telecommunications. Three levels of decisions are distinguished: design of structural elements of telecommunication networks, top level design of telecommunication networks and design of optimal policies of telecommunication enterprise. Examples of typical problems from each level show that the stochastic programming paradigm is a powerful approach for solving telecommunication design problems.

Keywords: stochastic optimization, telecommunications, network design, performance analysis, models of competition.

1 Introduction

Telecommunications have a long tradition of application of advanced mathematical modeling methods. Besides being a consumer of mathematical modeling, telecommunications provided a motivation for development of areas of applied mathematics. Important chapters of the theory of random processes have their roots in the work of telecommunication engineers. So far this mutual influence was mainly limited to the queueing theory and the theory of Markov processes, but now new decision problems arise which require the application of optimization methods. The recent trends in telecommunications have led to considerable increase in the level of uncertainty which became persistent and multifaceted. The decision support methodologies which provide adequate treatment of uncertainty are becoming particularly relevant for telecommunications. Stochastic optimization is the methodology of choice for optimal decision support under uncertainty, see Ermoliev and Wets (1988), Kall and Wallace (1994) Birge and Louveaux (1997). This paper provides a survey of applications of stochastic programming for solving design and decision problems in telecommunications.

Stochastic optimization is important for a large variety of such problems. We start by defining a classification which will serve as a roadmap for the exposition. This classification is made using the scale of the decision relative to the whole telecommunication environment which defines the nature of decision itself. Besides, different types of uncertainty come into play at different levels. We distinguish three scale levels: technological, network and enterprise. The technological level corresponds to the smallest scale and the enterprise level to the largest and the most aggregated scale.

The technological level deals with design of different elements of telecommunication networks, including switches, routers, multiplexers. Uncertainty on this level is a salient feature of commu-
nication requests and flows in the network. Besides, it can arise due to equipment failures. The key decisions are the engineering decisions which define the design for blueprints of these elements. Such blueprints depend on a number of parameters which should be chosen from the point of view of performance and quality of service. Traditionally, performance evaluation of the elements of telecommunication networks was the domain of queueing theory (Medhi, 1991). To be successful the methods of this theory require a specific probabilistic description of the stochastic processes which govern the behavior of communication flows. Usually such description is lacking for the new data services, and when it exists it does not satisfy requirements of the queueing theory. Stochastic optimization may help to obtain the performance estimates in the cases when more traditional methods are difficult to apply. See section 2 for one such example.

**Network level** problems deal with design and planning of different kinds of networks which may differ by scale and by technology involved. These can be access networks, local area networks, fixed or mobile networks. The decisions involve the placement of processing and link capacities provided by a given technology in a given geographic area with the aim to satisfy aggregated demand for telecommunication services from different user groups. Decisions are often of dynamic nature and include several time periods. The main uncertainty here is due to the demand for telecommunication services. Due to quantitative and qualitative explosion of such services, this kind of uncertainty increased considerably during the last decade. There are important additional sources of uncertainty connected with possible network failures and future technology development. Stochastic programming methods provide an added value of identifying the robust network design which within reasonable bounds will accommodate the future demand variations. This is particularly true for stochastic programming problems with recourse and multiperiod stochastic programming problems which provide intelligent means for mediation between different and often conflicting scenarios of the future. While traditional design approach is centered around minimization of the network costs under technological and quality of service constraints, systematic application of stochastic programming techniques includes incorporation of modern tools from corporate finance like evaluation of real options. Comprehensive models which include pricing decisions and binary variables provide a motivation for further development of this methodology. Section 3 contains several examples of stochastic optimization models for design problems at the network level. For related examples see Bonatti, Gaivoronski, Lemonche, and Polese (1994), Sen, Doverspike, and Cosares (1994), Fantauzzi, Gaivoronski, and Messina (1997), Tomasgard, Audestad, Dye, Stougie, der Vlerk, and Wallace (1998), Dempster and Medova (2001), Yen, Schaefer, and Smith (2001), Andrade, Lisser, Maculan, and Plateau (2002).

Finally, the **enterprise level** is the highest level of aggregation and looks at the telecommunication enterprise as a member of a larger industrial environment which includes other industrial actors and different consumer types. Decisions involve selection of the range of services which the enterprise will provide to the market, strategic investment decisions, pricing policy. Market acceptance of services, innovation process and actions of competition constitute the sources of uncertainty which are not present at the lower levels. Telecommunication and, more generally, information industry differs in important ways from traditional industries due to the rapid pace of innovation. This leads to the absence of perfect markets and to fundamental nonstationarity which makes it difficult to apply traditional microeconomic approaches based on equilibrium. Stochastic programming models enriched with selected notions of game theory can provide more adequate decision recommendations here. We outline one such model in section 4.

There is no rigid boundary between various levels since decisions taken at each level influence decisions on other levels. Further exposition is organized along the lines of this classification. Each
section contains examples of stochastic optimization models which illustrate the typical problems of each level. A summary concludes.

2 Technological level

The central problem is to find the design parameters of a piece of telecommunication equipment which will assure a given level of performance for a specified class of traffic patterns. We describe this problem on a general level and then provide a specific example which deals with access design of high speed data network.

In the vast majority of practically interesting cases the performance is measured by the functional

$$F(x, H) = \mathbb{E}_H f(x, \xi) = \int f(x, \xi) dH(\xi)$$  \hspace{1cm} (1)

where \(x\) is the vector of design parameters, \(\xi\) denotes the values of a stochastic process defined on an appropriate probability space which describes the interaction of traffic with device and \(H\) is the stationary distribution of this process. The function \(f(x, \xi)\) describes the performance of the equipment for a given traffic value \(\xi\) and a given value of design parameters \(x\). \(\mathbb{E}_H\) is the expected value with respect to \(H\). The values of the performance measure \(F\) should belong to the set \(\Phi\) of admissible values which describes requirements for the grade of service. This set is usually defined by bounds \(\phi^-\) and \(\phi^+\). In this case parameters \(x\) should satisfy

$$\phi^- \leq F(x, H) \leq \phi^+.$$  \hspace{1cm} (2)

Usually equipment should work satisfactorily for a sufficiently wide range of admissible traffic patterns which are defined by a set of traffic parameters. Therefore instead of a single distribution \(H\) in (1) we have a whole set of distributions \(G\) which is defined indirectly by traffic parameters of interest.

Design problem. Find the values of design parameters \(x\) for which the values of the performance functional \(F(x, H)\) belong to an admissible set \(\Phi\) for all \(H \in G\).

The final design is selected among solutions of this problem by considering additional criteria which may have economic or manufacturing nature.

The function \(f(x, \xi)\) from (1) is usually known and has a simple analytical structure. The major difficulty here is constituted by the distribution \(H\). The reason is that the traffic is described in terms of characteristics of individual nonhomogeneous traffic sources. Even if description of a single source is relatively simple, the composite traffic consisting of a large amount of such sources can be complex. Complexity increases due to nontrivial interaction of traffic with device. Even the description of a single source can be difficult to obtain, especially in the case of new services which generate traffic with partly unknown properties. Therefore a characterization of the distribution \(H\) and set \(G\) can be extremely difficult.

Traditionally, performance analysis developed two complementary approaches for confronting this difficulty: analytical analysis and simulation. The analytical approach requires that the traffic sources are described by a Markov chain. This is a serious limitation because the number of states in such a Markov chain can be high for realistic sources. After this the whole system consisting of the traffic and device is described by a Markov chain and its stationary distribution \(H\) is computed numerically. This enables a computation of the performance functional from (1) by direct integration which reduces to summation. The main problem of this approach is that the
resulting Markov chain is very often so large that the computation of its stationary distribution is far beyond the reach of modern computers. Approximations are necessary for a majority of the problems of interest which may undermine the relevance of results. The analytic approach is often supplemented by a simulation approach. It consists of direct simulation of the interaction between the traffic and device and allows a considerably more realistic representation of the whole system. The problem with this approach is that the simulation times necessary for obtaining the estimates of stationary performance can be prohibitively long. This is especially true for the case when design requirements are expressed in terms of packet loss which is a popular performance measure for the modern data networks. Both these techniques have the common drawback that they are applicable for a given traffic pattern, while design should be valid for the whole range of traffic parameters.

Stochastic optimization can enhance both analytic and simulation approaches to performance analysis by addressing the problem of development of guaranteed estimates for performance of telecommunication systems. Such estimates involve computation of $F^+(x)$ and $F^-(x)$ such that

$$F^-(x) \leq F(x, H) \leq F^+(x)$$

for all $H \in G$. When performance requirements are described by (2) one can select $x$ from

$$\phi^- \leq F^-(x), \ F^+(x) \leq \phi^+$$

which will guarantee satisfaction of performance requirements for all traffic patterns of interest. The bounds $F^+(x)$ and $F^-(x)$ can be obtained from the solution of

$$F^-(x) = \inf_{H \in G} F(x, H), \ F^+(x) = \sup_{H \in G} F(x, H).$$

These problems can be classified as belonging to a special class of stochastic optimization problems, namely optimization problems in the space of probability measures, see Karlin and Studden (1966), Kemperman (1968), Dupačová (1978), Ermoliev et al. (1986), Gaivoronski (1986). One may object that to solve such problems should be even more difficult than to compute the value of $F(x, H)$ for a given $H$, a difficult problem by itself as argued above. In many cases this is not true. Firstly, the function $F(x, H)$ can be often approximated by simpler functions $F^+(x, H)$ and $F^-(x, H)$ satisfying

$$F(x, H) \leq F^+(x, H), \ F^-(x, H) \leq F(x, H)$$

for all $H \in G$. These functions can be used in (4) instead of $F(x, H)$ which will make these problems simpler. Even more important, the measures which solve the problems (4) often have a very special structure which can be obtained from analysis of function $F(x, H)$ and set $G$ without solving the problem itself (Whitt, 1984; Ermoliev et al., 1986). Numerical methods have been developed which exploit this structure (Gaivoronski, 1986) and simplify the solution of (4).

### 2.1 Access engineering of broadband multiservice network

We illustrate this general approach using a specific example of access engineering of broadband multiservice network taken from Bonatti and Gaivoronski (1994b). We consider a high speed data network with data packets of fixed length, for example a network based on Asynchronous Transfer Mode (ATM) architecture, see De Prycker (1995). The central part of an access network consists of a server (multiplexer) with one output link and $L$ input links. Data packets arrive from input links and are sent by the server to the network, maybe staying some time in the buffer of limited capacity.
If some packet finds the buffer full upon arrival it is discarded. Since all the packets are of the same length, so are the service times. Therefore it is natural to consider the system operating in discrete time with the time interval being equal to the service time of one packet. Denoting by \( l(t) \) the buffer contents at time \( t \), by \( \xi_i(t) \) the number of packets which arrive from the source \( i \) at time \( t \) and by \( \xi(t) \) the total number of packets arrived at time \( t \) we have the following relation which describes the dynamics of the buffer contents:

\[
l(t + 1) = \max \{0, \min \{x_0, l(t) + \xi(t) - 1\}\}, \quad \xi(t) = \sum_{i=1}^{L} \xi_i(t),
\]

where \( \xi_i(t) \) is 0 or 1. Each source \( i \) at the input of a server generates a packet arrival process with distribution \( H_i \). This distribution defines the probability that a packet arrives at time \( t \) conditioned on the history of packet arrivals and is described by a vector of parameters \( a \):

\[
H_i = H_i(a).
\]

Many distributions may correspond to a given value of the vector \( a \). Each source belongs to one of \( N \) classes where each class corresponds to a given service. In terms of arrival distribution \( H_i \) each class is characterized by a set \( A_j \) in the space of parameters, such that if the source \( i \) belongs to the traffic class \( j \) it means that \( H_i \in G_j \) where

\[
G_j = \{H(a) \mid a \in A_j\}.
\]

Examples of traffic classes and corresponding parameter sets will be given later. Denote by \( x_j \) the number of sources which belong to class \( j \), then \( \sum_{j=1}^{N} x_j = L \). \( I_j \) is a subset of \( \{1, \ldots, L\} \) which indexes the sources belonging to the traffic class \( G_j \). The maximal admissible number of sources of each class \( x_j, j = 1 : N \) together with the buffer length \( x_0 \) constitute the vector \( x \) of design parameters which should be chosen so that the access system satisfies given performance requirements which are expressed in terms of the admissible upper bound \( \phi^+ \) on the packet loss probability \( F(x, H) \):

\[
F(x, H) = \mathbb{E} \max \{0, \xi(t) - l(t) - 1\} \leq \phi^+
\]

where \( H = H(z; x) = \mathbb{P} \{\xi(t) \leq z\} \) denotes the stationary distribution of \( \xi(t) \). Usually this bound is chosen between \( 10^{-10} - 10^{-9} \).

This and similar systems constitute an important part of telecommunication networks and considerable effort was dedicated to its study, see, for example, Labetoulle and Roberts (1994). The Markov chain analysis of this system starts by assuming that the packet arrivals from a single source are independent. This is a very serious assumption because the packets in ATM networks are generated by splitting larger amounts of data into small packets of the standard size. Therefore the traffic from a single source consists of bursts which correspond to each communication request. Even then, the resulting Markov chain contains at least \( \prod_{i=0}^{\infty} (x_i + 1) \) states which can be a very large number. The number of states explodes further if more realistic assumptions about the traffic are taken. A simulation approach also encounters difficulties because estimation of probabilities of the order of \( 10^{-10} \) requires long simulation times.

This problem can be treated by optimization over probability measures. We start by deriving a bound on the packet loss probability from (5).

**Proposition 1** (Bonatti and Gaivoronski, 1994b) Suppose that \( \xi_i(t) \) are the stationary ergodic stochastic processes for all \( i \) and the length of the buffer \( x_0 \) is not smaller than the total number of
sources \( L \). Then

\[
F(x, H) \leq F^+(x, H) = \frac{\max \{0, z - 1\} dH(z; x)}{zdH(z; x)} = \frac{E_H \max \{0, \zeta - 1\}}{E_H \zeta}
\]

where \( \zeta \) is a random variable distributed according to \( H \) and \( E_H \) is expectation with respect to \( H \).

The arrival processes from each source usually are assumed to be independent. Then the upper bound \( F^+(x) \) on the cell loss probability can be obtained by solution of

\[
F^+(x) = \max_{H_i \in G_j, \forall i \in I_j} \frac{1}{B(x, H)} \int \ldots \int \max \left\{ 0, \sum_{i=1}^{L} z_i - 1 \right\} \prod_{i=1}^{L} dH_i(z_i)
\]

where

\[
B(x, H) = \sum_{i=1}^{L} \int z_i dH_i(z_i).
\]

To advance further it is necessary to specify the sets \( G_j \) which define the traffic classes. One common way of doing so is to put bounds on the moments of the distributions belonging to this set together with the bounds on the support of these distributions. In this case

\[
G_j = G_j(a_j) = \left\{ H \mid \int_0^{a_{0j}} dH(z) = 1, \int_0^{a_{0j}} \psi_{ij}(z) dH(z) \leq a_{rj}, \ r = 1 : R \right\}
\]

where \( a_j = (a_{0j}, a_{1j}, \ldots, a_{Rj}) \) and \( \psi_{ij}(z) \) are known functions. From the point of view of access design this definition has an important meaning. The bound on support \( a_{0j} \) is the peak bandwidth of the source from the class \( j \) measured in fractions of the output bandwidth of the server. Let \( \psi_{1j}(z) = z \) then \( a_{1j} \) will be the average bandwidth of a source from the class \( j \). The properties of (7) with the sets \( G_j \) defined by (9) are well understood. Its solution has a special structure which was exploited for development of numerical methods in Ermoliev et al. (1986), Gaivoronski (1986).

Sometimes it is possible to obtain an explicit solution as in the important case when \( G_j(a_j) \) is defined by the values of peak and average bandwidth only:

**Theorem 2** (Bonatti and Gaivoronski, 1994b) Suppose that the traffic classes \( G_j \) are defined by the peak bandwidth \( a_{0j} \) and the average bandwidth \( a_{1j} \). Then among the sources which yield the largest packet loss probability always exist those with cell arrival distribution concentrated in two points \( (0, a_{0j}) \) with weights \( (1 - p_j, p_j) \), \( p_j = a_{1j}/a_{0j} \). The tight upper bound for the cell loss probability is

\[
F^+(x) = \frac{1}{\sum_{j=1}^{N} a_{1j} x_j} \sum_{0 \leq k_j \leq x_j} ^{\max} \left\{ 0, \sum_{j=1}^{N} a_{0j} k_j - 1 \right\} \prod_{j=1}^{N} \frac{x_j!}{k_j! (x_j - k_j)!} p_j^{k_j} (1 - p_j)^{x_j - k_j}.
\]

Design decision can be taken by finding feasible solutions of

\[
F^+(x) \leq \phi^+
\]
where $\phi^+$ is in Theorem 2. Design obtained by stochastic optimization has the following advantage compared to designs obtained by Markov chain modeling or simulations. It will assure the required quality of service for all sources with specified average and peak bandwidth. In contrast, the design obtained through Markov chain modeling will be valid only for a much narrower class of sources which in addition generate packets with independently distributed arrival times. Simulations can assure required quality of service only for a finite set of sources which were selected for simulation experiments.

Stochastic optimization can be applied to other design problems. Often it is important to exploit carefully the special structure of each particular case to obtain approximations of performance measure similar to (6). It is also possible to address the problem of this approximation from a more general point of view by developing approximations of complex random processes by simpler ones. For the case of Markov chains such approximations which yield guaranteed bounds for performance measures were developed in Bonatti and Gaivoronski (1994a).

3 Network level

The objective of the network level is to develop a design of the telecommunication network with a given capability to provide a set of services to a population of end users. Results of technological design are used as the inputs to the network design. This design should serve different and often conflicting purposes, e.g. satisfaction of demand, maintenance of a given service quality, cost effectiveness. Important decisions to take at this level include placement and dimensioning of processing nodes and transmission links. These decisions are affected by service pricing because it affects the quantity of demand to be satisfied. Exists considerable literature dedicated to the optimal design of networks in deterministic case (Ahuja et al., 1993; Bertsekas, 1998; Fortz et al., 2000; Mitra et al., 2001).

At the age of big state monopolies, largely immutable services and highly predictable environment the prevailing paradigm was the minimization of network costs under constraints on quality of service and demand satisfaction. Although this paradigm remains important, it is clearly insufficient for the highly mutable and uncertain environment of today. Robust network which can accommodate within reasonable bounds future market changes is more valuable then maybe cheaper network designed for today conditions and maybe for one specific future scenario. For this reason profit, service pricing, evaluation of investment opportunities under uncertainty become increasingly important in the network design models. This is where stochastic programming models have a competitive edge compared to deterministic models because the capability to mediate between scenarios of uncertain future is explicitly imbedded into them. This capability is very important because uncertainty of different kinds is one of the defining features of the modern telecommunications. On the network level the main source of uncertainty is unpredictable user response to introduction of new services which results in highly uncertain demand variations. To this one can add uncertainty due to technological innovation and uncertainty related to possible failures.

Another important feature of network design is the dynamic character of decisions. The network development projects have a time dimension and an important decision is how to distribute the investment over time. New information about the market will become available and the possibility to react on this information should be included in the decision models. Stochastic programming with recourse provides adequate tools for doing so. It also facilitates the incorporation of adaptation policies specifically designed to allow the network to react flexibly to the changing environment. Such policies are essential for robust network design. Another important issue is the correct evalua-
tion of flexibilities present in the network investment projects. Examples of such flexibilities or real options are option to expand, option to upgrade technology, option to alter usage. Consideration of these options can drastically change the overall evaluation of the network expansion project. For example a project which is unprofitable from the first glance can reveal hidden profit opportunities.

Further exposition is organized in the form of examples which illustrate the general ideas. Section 3.1 presents a series of decision models for planning of Internet based information service starting from a simple traditional deterministic cost minimization model to a two period stochastic programming model for profit maximization which can be used for evaluation of real options imbedded in this project. The problem of design of access network described in section 2 is considered in section 3.2 on the level of network design. This is an interesting example which shows interplay between both levels. Sections 3.3 and 3.4 show examples how technology influences network design models in the case of backbone networks. Network design which takes into account possible failures is considered in section 3.5. Finally, section 3.6 is dedicated to design of mobile networks in the situation of levelling out of demand.

3.1 Planning of Internet based information service

We consider here the problem of deployment of an Internet based information service on some territory which can be a country or a region. We assume that the network itself exists already and the decision consists in deployment of servers at the nodes of this network and assignment of demand generated in different geographical locations to these servers. The service provider on behalf of which the problem is solved can be the network owner, but can also be a virtual service provider which does not possess its own network and leases network from some network owner. Decisions to consider include phased introduction of service which can take the shape of Phase 1 deployment followed by Phase 2 deployment contingent on the market reaction. In addition, decisions include pricing of service.

Among various aspects of the problem one should consider geographical dimension, uncertainty of demand and costs, cost structure which includes fixed and variable costs, competition and substitution between services, relations between different market actors, e.g. network providers and service providers. The decision to go ahead with the project depends on the project profitability which in its turn depends on various options embedded in it, e.g. option to expand, to abandon, to upgrade technology. It is advisable to start the model development from the simple case which includes only some of the relevant features and to expand the model stepwise. We present here three such model development steps.

Step 1. Single period deterministic cost minimization model.

We start by considering only one decision period and full knowledge about demand and other uncertainties. Although these assumptions are highly unrealistic, the resulting model sets the stage for more realistic models. In this setting we assume that the deployment program has to satisfy the known demand fully. Since the service price is fixed, the revenue becomes fixed. For this reason the only way one can influence the profit is by minimizing the costs.

Notations

\( i = 1 : n \) - index for regions which constitute a territory. User population exists in each region which generates demand.

\( j = 1 : m \) - index for possible server locations.

\( y_j \) - binary variable which takes the value 1 if decision to place a server at location \( j \) is taken and 0 otherwise.
\( x_{ij} \) - amount of demand from region \( i \) served by server placed in location \( j \).

\( f_j \) - fixed cost for setting up a server in location \( j \).

\( c_{ij} \) - cost for serving of one unit of demand from region \( i \) by server at location \( j \).

\( d_i \) - demand generated at region \( i \).

\( g_j \) - capacity of server placed at location \( j \).

**Model**

Find the server deployment program \( y = (y_1, ..., y_m) \) and assignment of user groups to servers \( x = \{x_{ij}\}, i = 1 : n, j = 1 : m \) as solution of

\[
\min_{x,y} \sum_{j=1}^{m} f_j y_j + \sum_{j=1}^{m} \sum_{i=1}^{n} c_{ij} x_{ij} \tag{11}
\]

\[
\sum_{j=1}^{m} x_{ij} \geq d_i, \ i = 1 : n \tag{12}
\]

\[
\sum_{i=1}^{n} x_{ij} \leq g_j y_j, \ j = 1 : m \tag{13}
\]

where \( y_j \in \{0, 1\} \) and \( x_{ij} \geq 0 \). Here the first term in (11) represents the fixed cost of deployment of servers while the second term represents the variable costs for serving demand. The constraints (12) are imposed in order to obtain full demand satisfaction, while constraints (13) are the capacity constraints. This is a well known facility location model and it will serve as a starting point for development of a stochastic programming model with different scenarios of the future demand and a larger number of deployment phases.

**Step 2. Two period stochastic cost minimization model.**

Two deployment phases are considered: present Phase 1 with known demand and future Phase 2 with uncertain demand which is described by a finite number of scenarios. Each scenario is described by demand values in different regions during Phase 2 and the probability of this scenario. The Phase 2 decisions include additional deployment of servers and reassignment of demand to servers in response to the demand which becomes known. The model follows the framework of stochastic programming with recourse.

**Additional notations:**

\( r = 1 : R \) - index for demand scenarios.

\( d_i^r \) - demand generated by region \( i \) under scenario \( r \).

\( p_r \) - probability of scenario \( r \).

\( z_j^r \) - binary variable which takes the value 1 if under scenario \( r \) the decision to place a server at location \( j \) is taken and 0 otherwise.

\( x_{ij}^r \) - amount of demand from region \( i \) served by a server placed in location \( j \) under scenario \( r \).

\( \alpha \) - coefficient for discounting of the Phase 2 costs to the present.

Each scenario is characterized by a pair \((d^r, p^r)\) where \( d^r = (d_1^r, ..., d_n^r) \).

**Model:**

Find the Phase 1 server deployment program \( y = (y_1, ..., y_m) \), and assignment of user groups to servers \( x = \{x_{ij}\}, i = 1 : n, j = 1 : m \) as solution of

\[
\min_{x,y} \sum_{j=1}^{m} f_j y_j + \sum_{j=1}^{m} \sum_{i=1}^{n} c_{ij} x_{ij} + \alpha \sum_{r=1}^{R} p_r Q(r, y) \tag{14}
\]
subject to (12)-(13). The third term in (14) represents discounted costs of the Phase 2 deployment averaged over scenarios. The cost associated with scenario \( r \) is \( Q(r, y) \) and it depends on the Phase 1 deployment decision \( y \). These costs are obtained from solution of the recourse problem for each scenario \( r \):

\[
Q(r, y) = \min_{x^r, z^r} \sum_{j=1}^{m} f_j z_j^r + \sum_{j=1}^{m} \sum_{i=1}^{n} c_{ij} x^r_{ij}
\]

\[
\sum_{j=1}^{m} x^r_{ij} \geq d^r_i, \quad i = 1 : n
\]

\[
\sum_{i=1}^{n} x^r_{ij} \leq g_j (y_j + z_j^r), \quad j = 1 : m
\]

which is similar to (11)-(13) and chooses the Phase 2 deployment \( z^r = (z_1^r, \ldots, z_m^r) \) and new assignment of user groups to servers \( x^r = \{x^r_{ij}\}, i = 1 : n, j = 1 : m \) from minimization of fixed deployment costs and variable service costs for a given scenario \( r \).

This can be a numerically challenging problem because it contains binary variables. Still, the modern optimization technology permits to solve it for nontrivial and practically important cases. For example, we solved its deterministic equivalent to optimality using MPL modeling system powered by CPLEX and XPRESS solvers with \( R = 5, n = m = 20 \) in approximately 8 minutes on a 1133 Mhz Pentium III laptop. The deterministic equivalent for this case has 120 binary and 2400 continuous variables with 240 constraints. This time has grown to 1 hour for \( R = 6, n = m = 25 \) with the deterministic equivalent having 175 binary and 4375 continuous variables and 350 constraints. Utilization of decomposition is essential for solving the problems of larger dimensions.

Step 3. Two period stochastic profit maximization model with pricing.

In a competitive deregulated environment the profit maximization is a more appropriate objective than the minimization of network costs. It becomes fundamentally different from the plain cost minimization when the pricing decisions are considered simultaneously with deployment decisions. The models become more complicated because pricing affects demand and this dependence introduces nonlinearities. Still, meaningful analysis is feasible also in this case. We start by defining the linear demand model extending the scenario framework explained before.

Additional notations:
\( h_0 \) - reference price for service during Phase 1.
\( d_{i0} \) - reference demand at region \( i \) during Phase 1 which corresponds to reference price \( h_0 \).
\( w_i \) - demand elasticity at region \( i \) during Phase 1.
\( h \) - the price increment relative to the reference price during Phase 1.
\( h^r_0 \) - reference price for service during Phase 2 under scenario \( r \).
\( d_{i0}^r \) - reference demand at region \( i \) during Phase 2 which corresponds to reference price \( h_0^r \) under scenario \( r \).
\( w_i^r \) - demand elasticity at region \( i \) during Phase 2 under scenario \( r \).
\( h^r \) - the price increment relative to the reference price during Phase 2 under scenario \( r \).

Demand model.

This is the crucial piece of the profit model. Consider the Phase 1 deployment. The service price equals \( h_0 + h \) and the price decision consists in selecting the price increment \( h \) which may be positive or negative. Assume that the demand \( d_i \) in region \( i \) during this phase depends on the price
of the service according to some function \( d_i = f_i(h_0 + h) \) and in the vicinity of the point \( h_0 \) this dependence can be linearized via
\[
d_i = d_{i0} - w_i h. \tag{18}
\]
Similar relations describe the demand behavior during Phase 2 for each of the scenarios \( r = 1 : R \). Each scenario is defined in this case by a tuple \((d_{i0}^r, h_0^r, w_i^r, p^r)\) which defines the dependence of demand on price according to relation (18) for a given scenario \( r \).

**Decision model.**

Find the Phase 1 increment for the service price \( h \), server deployment program \( y = (y_1, ..., y_m) \), and assignment of user groups to servers \( x = \{ x_{ij} \}, i = 1 : n, j = 1 : m \) as a solution of

\[
\max_{h,x,y} W(h) = C(y, x) + \alpha \sum_{r=1}^{R} p_r P(r, y) \tag{19}
\]
subject to

\[
w_i h + \sum_{j=1}^{m} x_{ij} \geq d_{i0}, \ i = 1 : n \tag{20}
\]
and constraint (13). Here \( W(h) \) is the revenue during Phase 1:
\[
W(h) = \sum_{i=1}^{n} (h + h_i) (d_i - w_i h) \tag{21}
\]
and \( C(y,x) \) are the costs during Phase 1 defined according to (11). The third term in (19) represents the profits during Phase 2 averaged over scenarios and discounted to the present where \( P(r, y) \) is the profit during Phase 2 under scenario \( r \). It is taken as the optimal value of the following recourse problem:
\[
P(r, y) = \max_{h',x',x''} W(r, h') - C(r, z', x') \tag{22}
\]
subject to additional constraints (17). Here \( W(r, h') \) is the revenue during Phase 2 under scenario \( r \) obtained similarly to (21) and \( C(r, z', x') \) are the costs during Phase 2 under scenario \( r \) taken from (15).

There is one important feature of this model which was not present in the models (11)-(13) and (14)-(17). While (14)-(17) can be transformed to a mixed integer LP by considering the deterministic equivalent, no such transformation is possible for the model (19)-(23). This is because the revenues \( W(r, h) \) and \( W(r, h') \) depend nonlinearly on the decision variables \( h \) and \( h' \). Even in the simplest case of the linear demand model (18) this dependence is quadratic. Therefore specialized numerical techniques should be employed in this case with decomposition approaches being the most promising.

**Evaluation of investment opportunities, real options.**

One of the most important utilizations of model (19)-(23) is the evaluation of profitability of the investment project which consists in the deployment of the new service. Recent developments in the corporate finance showed importance of evaluation of real options for correct evaluation of industrial projects (Trigeorgis 1996). While for more traditional industries direct evaluation
techniques can be similar to evaluation of financial options, for innovative industries with unique projects such approaches are difficult to apply. Stochastic programming models can represent a valid alternative for real option evaluation. Let us utilize model (19)-(23) for this purpose. In particular, let us evaluate options to expand, to upgrade technology, to abandon or to convert a part of the infrastructure.

Option to expand (wait and see option).

This option is already imbedded in the model (19)-(23) which contains the possibility to add additional servers during Phase 2 contingent on market reaction. The value of this option can be computed as follows. Denote by $P^*$ the optimal value of the model (19)-(23). This is the value of the project with an option to expand. The value $\bar{P}$ of the same project without the option to expand is obtained by solving the same model with binary variables $z^r_j$ fixed to zero for all scenarios. Clearly, $\bar{P} \leq P^*$. The value of the option is the difference $P^* - \bar{P}$.

Option to upgrade technology.

This is a valuable option because it can dramatically change the project evaluation. The most known example is the GSM mobile network whose development started when the technology to make mobile phones small enough was not yet available. In order to evaluate this option it is necessary to have a closer look at the ways the technology development can affect various components of the model (19)-(23). For example, the technology development can lead to decreasing of fixed costs for server installation and/or increase in the possible server capacities during Phase 2. In this case it is necessary to introduce these features into definition of scenarios. The notations are

- $f^r_j$ - fixed cost for setting up server in location $j$ under scenario $r$.
- $g^r_j$ - capacity of server placed at location $i$ under scenario $r$.

The model changes as follows. The part (19)-(20) remains the same because it describes Phase 1 to be implemented with known technology. The part (22)-(23) has a modified capacity constraint which substitutes (17):

$$\sum_{i=1}^{n} x^r_{ij} \leq g^r_j y^r_j + g^r_j z^r_j, \quad j = 1 : m.$$  \hspace{1cm} (24)

Besides, the cost term $C(r, z^r, x^r)$ from (22) is

$$C(r, z^r, x^r) = \min_{x^r, z^r} \sum_{j=1}^{m} f^r_j z^r_j + \sum_{i=1}^{n} \sum_{j=1}^{m} c^r_{ij} x^r_{ij}$$  \hspace{1cm} (25)

The model (19)-(23) is solved with modification (24)-(25) which will give the value $P^{**}$ of the project with an option to upgrade technology. This value is compared with the value of the project $P^*$ without such an option and the difference $P^{**} - P^*$ will give the option value.

Option to abandon.

This is a valuable option when the market reaction is uncertain. If demand does not catch up it is reasonable to cut maintenance costs in the regions where demand is weak and possibly recover part of the fixed costs by selling or leasing the server infrastructure. The additional notations are:

- $b^r_j$ - maintenance costs for server at location $j$ during Phase 2 under scenario $r$.
- $\beta^r_j$ - fraction of fixed costs which can be recovered by abandonment of server at location $j$ under scenario $r$.
- $u^r_j$ - binary variable which equals 1 if server at location $j$ is abandoned during scenario $r$.

The model changes as follows. Again, the part (19)-(20) which refers to the Phase 1 remains the same. The part (22)-(23) has a modified capacity constraint which substitutes (17):
\[
\sum_{i=1}^{n} x_{ij}^r \leq g_j (y_j + z_j^r - u_j^r), \ j = 1 : m, \ r = 1 : R
\] (26)

and additional abandonment constraints
\[
u_j^r \leq y_j, \ j = 1 : m, \ r = 1 : R.
\] (27)

The revenue term \(W(r, h^r)\) and the cost term \(C(r, z^r, x^r)\) from (19) are
\[
W(r, h^r) = \sum_{i=1}^{n} (h_0^r + h^r_i)(d_{i0}^r - w_i^r h^r) + \sum_{j=1}^{m} \beta_j^r f_j u_j^r
\] (28)

\[
C(r, z^r, x^r) = \sum_{j=1}^{m} \left( f_j z_j^r + b_j^r (y_j + z_j^r - u_j^r) + \sum_{i=1}^{n} c_{ij} x_{ij}^r \right).
\] (29)

Figure 1: Evaluation of real options in the case of service introduction

The model (19)-(23) is solved with the modification (26)-(29) which will yield the value \(P^{++}\) of the project with an option to abandon infrastructure. This value is compared with the value of the project \(P^+\) obtained by solving the same model with variables \(u_j^r\) set to zero which corresponds to evaluation without the option to abandon. The difference \(P^{++} - P^+\) is the option value.

We now provide an example of such an option evaluation depicted in Figure 1. This figure shows the dependence of the project value on the service price \(h_0 + h\) for the case when the Phase 2 service prices were fixed to the Phase 1 prices, i.e. \(h_0^r \equiv h_0, h^r \equiv h\). Three alternatives are shown in this figure. The first alternative is depicted by a thin curve and describes the dependence of project value on price for the case when no option to expand and no option to upgrade technology are
considered during Phase 2. The second alternative allows an option to expand, but not an option to upgrade technology and is depicted by a dotted curve. The third alternative shown with a thick line allows both options during Phase 2.

First of all, one notices the jumps on the curves which are due to the discrete character of the decisions. The objective in all three cases is full demand satisfaction. A small increase in price leads to a small decrease in demand which can make a given server redundant with a corresponding stepwise decrease in fixed costs. Another observation confirms the added value of flexibility which options provide. The value of the project without option is not positive even for the best choice of service price. The project becomes profitable when the option to expand is allowed. There are two regions of profitability with respect to the service price. The first corresponds to an aggressively low service price designed to stimulate large demand and the second corresponds to a less aggressive behavior with high prices and smaller demand. These profitability regions expand when an additional option to upgrade technology is considered. In the absence of options the model recommends defensive behavior with high pricing, while flexibility imbedded in options allows to stimulate demand more aggressively with lower prices.

3.2 Design of access network

We consider the topic of section 2.1 from the network level. The model from the technological level allowed us to understand which user population can have a guaranteed quality of service from an access server with given technological characteristics. On the basis of this information the network level design should answer the following questions. Given the present demand and future projections for a given region, what type of servers and how many of them a network operator should choose? Given a description of the typical user populations with given demand patterns, what kind of access servers should the equipment manufacturer produce? In this section we describe a stochastic programming model which helps to answer these questions (Bonatti et al., 1994).

We focus on the core of the network access design which is the design of the first statistical multiplexer stage. The model will support decisions on the number of multiplexers, the bandwidth they carry, a recommended composition of the traffic they serve. Such a model with two time periods: "the present" and "the future" incorporates some robustness against unexpected traffic evolution, otherwise the design can become obsolete less or more rapidly and considerable problems due to the system reconfigurations can arise. The present demand is relatively well known. The information available at present about the future demand is available in the form of several demand scenarios. The decision about access design should be made at the "present" only on the basis of scenario probabilities, without knowledge about which of the demand scenarios will materialize. When the demand will become known in the future some corrective action will be taken with the aim to assure the required grade of service. The objective is to select a design which provides a cost effective solution both from the point of view of the "present" costs (design implementation) and the "future" costs (design correction).

Evolution of traffic patterns

- Traffics generated by different users are statistically independent.
- Traffic generated by any of the users belong to one of the well defined traffic classes whose characteristics are known. These classes are defined by simple parameters like maximal and average bandwidth. Initially there are \( N \) traffic classes \( \nu_i, i = 1 : N \), in the future there are \( M \) traffic classes \( \mu_j, j = 1 : M \). Some future classes may coincide with the past classes, while some others may be completely new.
- Users, presently belonging to class \( \nu_i \) can pass in the future to class \( \mu_j \), new users can appear in class \( \mu_j \).

- Present traffic classes are known; for the future traffic classes there are \( K \) different scenarios \( \theta_k \), \( k = 1 : K \) described as triples

\[
\theta_k = \{ X^k_0, \alpha^k_{ij}, \rho^k \} , k \leq K, i = 0 : N, j = 1 : M, \sum_{j=1}^{M} \alpha^k_{ij} = 1, \sum_{j=1}^{M} \rho^k = 1
\]

where \( X^k_0 \) is the number of new users under scenario \( \theta_k \), \( \alpha^k_{ij} \geq 0 \) for \( i = 1 : N \) is the portion of the users of class \( \nu_i \) which pass to class \( \mu_j \) under scenario \( \theta_k \) and for \( i = 0 \) it is the portion of the new users which belong to the class \( \mu_j \), \( \rho^k \geq 0 \) is the probability of scenario \( \theta_k \).

Denoting the number of users which belong to the present traffic class \( \nu_i \) by \( X_i, i = 1 : N \) and the number of users which belong to the future traffic class \( \mu_j \) under scenario \( \theta_k \) by \( Y^k_j \) we obtain the following relation between these two quantities:

\[
\sum_{i=1}^{N} \alpha^k_{ij} X_i + \alpha^k_{ij} X^k_0 = Y^k_j, \quad j = 1 : M.
\]

Decision variables

We consider the case when all multiplexers from the first statistical multiplexing stage handle a fixed mixture of sources. The more general case which includes multiplexers of different types can be handled similarly. Thus, we have to define the initial number \( n_1 \) of multiplexers to install, the number of multiplexers \( n_2 \) to add later in the future when the demand scenario \( \theta_k \) becomes known, the bandwidth \( a \) of one multiplexer and the mix of sources served by an arbitrary multiplexer at present and in the future. \( x_i \) is the number of users of class \( \nu_i \) supported by one multiplexer at present and \( y^k_j \) is the number of users of class \( \mu_j \) supported by one multiplexer in the future under scenario \( \theta_k \). The vectors \( x = (x_1, \ldots, x_N) \) and \( y^k = (y^k_1, \ldots, y^k_M) \) describe the source mix. These decision parameters satisfy the quality constraints of two types. The first one states that all users should be covered, now and in the future:

\[
n_1 x_i \geq X_i, i = 1 : N
\]

\[
(n_1 + n_2 y^k_j) \geq Y^k_j, \quad j = 1 : M, k = 1 : K
\]

where \( Y^k_j \) is defined in (30). The second group of constraints should assure that each multiplexer provides the required grade of service. The quality of service (QoS) is characterized by the known function \( f(a, z) \) where \( a \) is the multiplexer bandwidth and \( z \) is the source mix supported by a given multiplexer. There is an admissible bound \( \gamma \) on the grade of service. This yields the following representation for the QoS constraints for the present and the future:

\[
f(a, x) \leq \gamma
\]

\[
f(a, y^r) \leq \gamma, \quad r = 1 : R.
\]

The specific expression for the function \( f(a, z) \) or the algorithm for its computation is provided by the technological level design where an example was described in section 2.1. If we take the packet loss as the measure of the quality of service we can use expression (10) from that section.

Costs
We take into account the costs of initial installation, additional installation, connection of new users, reconnection of old users.

- cost of initial installation and connection of users:

\[ n_1 C_{11} + C_{21} \sum_{i=1}^{N} X_i \]

where \( C_{11} \) is the fixed cost for initial installation of one multiplexer and \( C_{21} \) is the initial cost for connecting one user;

- cost of additional installation and connection of users:

\[ n_2 C_{12} + C_{22} X_0^k \]

where \( C_{12} \) is the fixed cost for additional installation of one multiplexer in the future and \( C_{22} \) is the cost for connecting one user in the future;

- cost of reconnecting the users in the future:

\[ C_3 n_1 \sum_{j=1}^{M} \max \left\{ 0, \sum_{i=1}^{N} \alpha_{ij} x_i - y_i^k \right\} \]

where \( C_3 \) is the cost of reconnecting of one user in the future. The future costs \( C_{12}, C_{22} \) and \( C_3 \) may be only partially known and may differ between different scenarios.

**Decision hierarchy.**

Different decision variables are defined by two coordinated optimization problems. They correspond to different time scales and different levels of knowledge about demand. Initial installed capacity \( n_1 \), multiplexer bandwidth \( a \) and initial source mix \( x \) are decided at the present, when the actual future demand scenario is not known. This is the **long term planning problem**. Additional installed capacity \( n_2^k \) and the final source mix \( y^k \) supported by one multiplexer are decided in the future when the demand scenario is known. This is the decision correction problem or **recourse problem** in stochastic programming terminology. The optimal value of this problem enters in the expression for the total costs of the long term planning problem. This recourse problem is defined as follows.

For scenario \( k \), multiplexer bandwidth \( a \), initial installed capacity \( n_1 \) and initial traffic mix \( x \) find \((n_2^k, y^k)\) which solve

\[ Q(k, a, n_1, x) = \min \left( n_2^k, y^k \right) \left( C_{12} n_2 + C_{22} X_0^k + C_3 n_1 \sum_{j=1}^{M} \max \left\{ 0, \sum_{i=1}^{N} \alpha_{ij} x_i - y_i^k \right\} \right) \] (35)

subject to constraints (32),(34).

The stochastic programming problem with recourse for long term planning is to find \((a, n_1, x)\) which minimize the present costs and discounted future costs averaged among demand scenarios:

\[ \min_{a, n_1, x} C_{11} n_1 + \alpha \sum_{k=1}^{K} p^k Q(k, a, n_1, x) \] (36)

subject to constraints (31),(33) where \( \alpha \) is a discount coefficient.
It possesses features which place it apart from the vast majority of such problems found in the literature. First of all, constraints (33),(34) are nonlinear. Moreover, while variables $a, x, y^k$ can be considered to be continuous, the variables $n_1$ and $n_2$ are substantially discrete. Therefore the usual solution approaches based on application of large scale linear programming to the deterministic equivalent or even Benders decomposition are not applicable here. A version of stochastic random search worked well on this problem.

3.3 Design of backbone connection oriented network

The objective of the network design is to make a decision about the placement of the network elements in a given geographical area. The network is represented as the collection of nodes with processing capabilities which are connected by links with transmission capacities. Design involves taking decisions concerning the placement of nodes of different type and processing capability and placement of links with different capacities. The objective of design is the satisfaction of communication demand between different nodes under given requirements about the quality of service and taking into account profit, costs and other considerations. Telecommunication networks have a hierarchical structure and backbone networks are the top level of this hierarchy responsible for carrying large demand quantities between geographically distributed nodes where demand is collected with the help of local networks.

The design of backbone network involves considerable investments and the resulting network should be robust enough to accommodate unpredictable demand variations during the time horizon of a few years. The flexibility required for the adequate reaction to the changing demand patterns is provided by the network management policies. In the case of connection oriented networks such policies may include construction of a logical network on the top of the physical network by reserving transportation capacity between different nodes for different virtual paths in the network. While the change in the physical network requires considerable investment and time, the change in the logical network can be performed relatively fast and cheap following the change of the demand pattern. The design of the physical network should take into account this possibility and the stochastic programming approach provides the necessary modeling tools for doing so. We describe one such model for the case of connection oriented networks where the transportation and processing capacity for connection between any given pair of nodes should be reserved before the actual communication can take place. Examples of such networks are traditional telephone networks, broadband ATM networks (De Prycker, 1995), mobile networks.

Topology of the physical network.

- $n$ - number of nodes in the network. The nodes are indexed by integer numbers $i = 1 : n$.
- $x_{ij}$ - link capacity between nodes $i$ and $j$.
- $u_i$ - processing capacity of node $i$.
- $x$ - vector of all link capacities.
- $u$ - vector of all processing capacities.
- $x^-_{ij}$ - lower bound on the transmission capacity between nodes $i$ and $j$. Nonzero $x^-_{ij}$ means that the link between nodes $i$ and $j$ exists already and the objective of design is to identify the necessary network expansion.
- $x^+_{ij}$ - upper bound on the transmission capacity between nodes $i$ and $j$.
- $u^+_{i}$ - upper bound on processing capacity at node $i$.
- $u^-_{i}$ - lower bound on processing capacity at node $i$.

Demand scenarios.
$q = 1, \ldots, m$ index for demand scenarios.

- $d_{ij}^q$ - communication demand between nodes $i$ and $j$ under scenario $q$.
- $z_{ij}^q$ - amount of demand between nodes $i$ and $j$ which is not served under scenario $q$.
- $g^q$ - amount of the link capacity required for transportation of one demand unit by an arbitrary link under scenario $q$.
- $h^q$ - amount of communication flow through a node served by one unit of processing capacity at this node under scenario $q$. The communication flow is measured by the amount of transmission capacity necessary to carry this flow.
- $p_q$ - probability of scenario $q$.

The dependence of $g^q$ and $h^q$ on the demand scenario allows different service development possibilities.

**Topology of the logical network.**

Capacities of the links in the logical network depend on the specific demand scenario because this network should be adapted to demand.

- $I_{ij}$ - the set of admissible paths between nodes $i$ and $j$.
- $m_{ij}$ - number of different paths in the set $I_{ij}$.
- $b_{ijr}$ - path number $r$ from the set $I_{ij}$, $r = 1 : m_{ij}$. It is represented as a sequence of pairs of nodes where each pair identifies the link which belongs to the path $b_{ijr}$.

$$b_{ijr} = ((i, j_{r1}), (j_{r1}, j_{r2}), \ldots, (j_{rt_r}, j))$$

where $t_r$ is the number of hops in the path $b_{ijr}$.

- $y_{ijr}^q$ - capacity of path $b_{ijr}$ between nodes $i$ and $j$ under scenario $q$.

**Costs.**

The decision paradigm of the minimization of the network costs under constraints on quality of service and demand satisfaction is adopted here. Different objectives like maximization of revenue or maximization of profit can be considered.

- $c_{ij}^l$ - variable cost of installation of one unit of link capacity between nodes $i$ and $j$.
- $c_{ij}^f$ - fixed cost of installation of link capacity between nodes $i$ and $j$.
- $c_i^l$ - variable cost of installation of one unit of processing capacity in node $i$.
- $c_i^f$ - fixed cost of installation of processing capacity in node $i$.
- $e_{ij}^q$ - opportunity cost of not meeting one unit of demand between nodes $i$ and $j$ under scenario $q$.

**Decision structure.**

There are two coordinated decision problems here which fit the paradigm of stochastic programming problems with recourse. The decision to construct the physical network is taken now. Besides decisions about the specific values of transmission capacities $x_{ij}$ and processing capacities $u_i$, this decision involves also the logical decisions to install capacities or not. Due to the presence of transaction costs these decisions should be modeled by binary variables.

- $v_{ij}^q$ - binary variable which equals 1 if the link capacity between nodes $i$ and $j$ is increased above the already existing level $x_{ij}^-$ and zero otherwise.
- $w_i^q$ - binary variable which equals 1 if the processing capacity at node $i$ is increased above the already existing level $u_i^-$ and zero otherwise.

The design of the physical network is taken before the actual demand patterns become known only on the basis of the information about demand scenarios. Therefore the objective is to minimize the current costs of network installation and discounted future costs of the network adaptation to
demand, averaged among demand scenarios. The design problem is to find \( v_{ij}, w_i, x_{ij}, u_i \) for all \( i, j \) that solve
\[
\min_{v_{ij}, w_i, x_{ij}, u_i} \sum_{i,j} \left( c_i v_{ij} + c_{ij} x_{ij} \right) + \sum_i \left( c_i^r w_i + c_i^s u_i \right) + \alpha \sum_{q=1}^m p_q Q(q, x, u) \tag{37}
\]
\[
x_{ij}^+ \leq x_{ij} \leq x_{ij}^- + (x_{ij}^+ - x_{ij}^-) v_{ij}, \; \forall (i, j) \tag{38}
\]
\[
u_i^- \leq u_i \leq u_i^+ + (u_i^+ - u_i^-) w_i, \; i = 1 : n \tag{39}
\]
The objective function (37) includes fixed and variable costs for installation of processing capacities at nodes, transmission capacities at links and averaged costs of the network adaptation to demand discounted with the discount coefficient \( \alpha \). Constraints (38), (39) impose bounds on capacities and connect logical and continuous decision variables.

The cost \( Q(q, x, u) \) of the network adaptation to a given demand pattern \( q \) is obtained by solving the recourse problem which is the design problem of the logical network for this demand pattern. This problem will be solved repeatedly during the life time of the physical network as a new demand scenario emerges.

Given capacities \( x_{ij}, u_i \) of the physical network and demand scenario \( q \) find the link capacities \( y_{ijr}^q \) of the logical network by solving
\[
Q(q, x, u) = \min_{y_{ijr}, z_{ij}} \sum_{i,j=1}^n e_{ij}^q z_{ij}^q \tag{40}
\]
\[
\sum_{r \in I_{ij}} y_{ijr}^q + g^q z_{ij}^q = g^q d_{ij}^q, \; \forall (i, j) \tag{41}
\]
\[
\sum_{r \in I_{ij}, (k,l) \in b_{ijr}} y_{ijr}^q \leq x_{kl}, \; \forall (k, l) \tag{42}
\]
\[
\sum_{r \in I_{ij}, (k,l) \in b_{ijr}} y_{ijr}^q \leq h^q u_k, \; k = 1 : n \tag{43}
\]
\[
z_{ij}^q \geq 0, \; y_{ijr}^q \geq 0. \tag{44}
\]
The objective function (40) represents adaptation costs which consist of the opportunity costs of not meeting demand. The costs of reconfiguring the logical network are assumed to be either negligible or not dependent on the decision variables. Constraint (41) connects the link capacities of the logical network with the served demand. In the left hand side of constraint (42) we have the total communication capacity required by the logical network from the link between nodes \( k \) and \( l \) which should not exceed the physical capacity \( x_{kl} \). The left hand side of constraint (43) represents the sum of all ingoing and outgoing communication flow at node \( k \) measured by reserved transmission capacity. It is assumed that the required processing capacity at node \( k \) is proportional to this flow.

Solving the problem (37)-(39) is via formulating its deterministic equivalent and solving the resulting mixed integer LP. Decomposition techniques may be obligatory to process problems of realistic dimensions. Possible simplifications include approximation of fixed costs by variable costs which allows to dispense with binary variables. When the bottleneck is represented by either processing capacities or transmission capacities the problem can be simplified by considering only the bottleneck capacities. Different variants of problem (37)-(39) were considered in Fantauzzi et al. (1997) where also the results of numerical experiments were reported.
3.4 Design of backbone connectionless network

Connectionless networks represent an important class of telecom networks where no capacity reservation is needed in order to establish communication between different nodes. An important example is the Internet network. The communication between nodes consists of the flow of data packets which are routed with the help of routing tables and routing algorithms implemented at network nodes. Therefore the logical network can be represented in the form of multicommodity network flow where each commodity corresponds to a given pair of nodes. The design objectives and the design paradigm remain the same as in the case of the connection oriented network, the main difference being in the formulation of the network adaptation problem (40)-(44).

The topology of logical network is represented by $y_{ijkl}^q$ which is the communication flow between nodes $i$ and $j$ which passes through link between nodes $l$ and $k$ under scenario $q$: The problem of design of the physical network (37)-(39) remains the same, while in the network adaptation problem (40)-(44) constraints (41)-(44) are substituted by the following constraints:

\[
\sum_{l:l\neq k} y_{ijkl}^q - \sum_{l:l\neq k} y_{ijlk}^q = 0, \forall (i, j), \forall k : k \neq i, j, \tag{45}
\]

\[
\sum_{l:l\neq i} y_{ijil}^q = \sum_{l:l\neq i} y_{ijli}^q + g^q z_{ij}^q = g^q d_{ij}^q, \forall (i, j) \tag{46}
\]

\[
\sum_{l:l\neq j} y_{ijjl}^q - \sum_{l:l\neq j} y_{ijlj}^q = -g^q d_{ij}^q, \forall (i, j) \tag{47}
\]

\[
\sum_{(i,j)} y_{ijkl}^q + \sum_{(i,j)} y_{ijlk}^q \leq x_{kl}, \forall (k, l) \tag{48}
\]

\[
\sum_{(i,j), l:l\neq k} y_{ijkl}^q + \sum_{(i,j), l:l\neq k} y_{ijlk}^q \leq h^q u_k, k = 1 : n \tag{49}
\]

\[
z_{ij}^q \geq 0, y_{ijkl}^q \geq 0. \tag{50}
\]

Constraints (45)-(47) are the network flow continuity constraints. In the left hand side of constraint (48) we have the total communication capacity required by the logical network from the link between nodes $k$ and $l$ which should not exceed the physical capacity $x_{kl}$. The left hand side of constraint (49) represents the sum of all ingoing and outgoing communication flow at node $k$. It is assumed that the required processing capacity at node $k$ is proportional to this flow.

3.5 Incorporating reliability considerations

An important source of uncertainty inherent in the telecommunication networks is presented by possible failures of links and nodes which can occur due to a variety of reasons, see Gavish and Neuman (1992). Therefore reliability and dependability of networks is a serious design issue. We show that the reliability considerations can be naturally incorporated into the stochastic programming modeling approach. We show how to extend the models of sections 3.3 and 3.4 to obtain a reliable network design.

The key is to represent the possible link and node failures in the form of failure scenarios which are similar to demand scenarios described in the previous sections. Such scenarios can be
independent from demand scenarios or can be combined with them. Demand scenario $q$ is described by the following quantities.

- $\alpha_{ij}^q$ - portion of the link capacity between nodes $i$ and $j$ which remains operational under failure scenario $q$, $0 \leq \alpha_{ij}^q \leq 1$.
- $\beta_k^q$ - portion of the processing capacity at node $k$ which remains operational under failure scenario $q$, $0 \leq \beta_k^q \leq 1$.
- $p_q$ - probability of scenario $q$.

Among all the failure scenarios there will always be one scenario $q = 1$ corresponding to normal operation when $\alpha_{ij}^1 = 1, \beta_k^1 = 1$ for all links $(i,j)$ and all nodes $k$. All other scenarios will correspond usually to the failure of one given link or node because in the vast majority of practical situations the simultaneous failure of several links or nodes is unlikely.

The models from sections 3.3 and 3.4 remain the same the only difference being a slight modification of capacity constraints. For connectionless network constraints (48),(49) are substituted by

$$\sum_{(i,j)} y^q_{ijkl} + \sum_{(i,j)} y^q_{ijlk} \leq \alpha_k^q x_{kl}, \forall (k,l)$$

(51)

For connection oriented networks similar modifications should be made to constraints (42),(43).

The discussion of this and the previous two sections by presenting the Figure 2 which shows dependence of the optimal network costs on the opportunity costs of not meeting demand. The specific model for which these costs were calculated was the design of a connectionless reliable network. The opportunity costs $e_{ij}^q$ are the same for all links and all scenarios.

The thin curve on the Figure 2 represents the dependence of the total network costs on the unit opportunity costs $e$. These costs equal the optimal solution of the problem (37)-(39) where $Q(q,x,u)$ is the optimal solution of the problem (40) with constraints (45)-(47),(50),(51),(52). These costs grow linearly for low opportunity costs $e$, after some threshold they start to grow more slowly and after another threshold they stop to grow which means that all demand is satisfied after that threshold. The dependence of the total costs on opportunity costs is concave. The thick curve represents the corresponding dependence of the pure network component of costs, namely the sum of the first two terms in (37). It shows some interesting phenomena which can be observed also in the common practice of the network development. When the unit opportunity costs of not meeting demand are low, no network is built at all and all the costs consist of penalties for not meeting demand. After the unit opportunity cost reach some threshold the network is built which satisfies a large portion of demand and this network remains unchanged with further growth of opportunity costs until another threshold is reached. After that second threshold the network is built which satisfies all demand. An important observation is that the network is upgraded not incrementally, but in a few large steps which corresponds to the industrial practice of the telecom network deployment. Stochastic programming models coupled with the estimation of opportunity costs can help to decide about the timing of decisions to upgrade the network.

### 3.6 Planning of capacity expansion of mobile network

The recent few years were characterized by exponential growth of mobile networks driven by exponential growth of demand. The market was capable of absorbing practically any capacity which
mobile operators were capable to offer. The things are changing now with the market for voice services being close to saturation in many countries while the market for the future broadband wireless services is uncertain. The mobile operators are becoming more attentive to the optimal network planning. The objective is to avoid potential losses which may be caused from one side by deployment of excessive capacity and from the other side by deterioration of quality of service and the consequent loss of customers who may turn to competition. Stochastic programming models can provide the network designers with balanced and robust network expansion alternatives which take into account uncertainties in demand development.

In an example the objective is to develop a plan for the expansion of mobile network based on the demand forecast. The time horizon of the network expansion is 1-2 years while the plan itself is revised periodically, for example each quarter. The expansion plan consists of two components: establishing new channels (TRX) in existing cells and establishing new cells. Establishing of new channels is a relatively minor matter compared to establishing of a new cell both from the point of view of expenditure and time. Establishing of a cell is a complex process which may take up to six months of time and considerable amount of resources. It consists of the following steps.

- Preparatory period during which the necessary agreements are made, it takes approximately half the time and a few percent of resources.
- Building of a cell which takes one third of the time and two fifth of resources.
- Putting the cell into operation which takes one sixth of the time and three fifths of resources.

It is possible to suspend the establishment of a cell after each of these steps and resume it again after some time. This contains a source of additional flexibility which is possible to exploit in order to diminish the reaction time on demand changes and avoid unnecessary investment. In particular, it is possible to have pools of semi-finished cells on different levels of preparedness and to invest into

Figure 2: Dependence of the optimal network costs on the opportunity costs of not meeting demand
further establishment only when demand requires this. The model which we present below helps to exploit such a possibility. The network expansion process is extended over several time periods \( t = 1, 2, ..., T \) which in this setting can have the length of one month, while \( T \) may take the values between 6 and 18.

**Demand description.** It is the key input in the network expansion model. As in the previous sections, demand forecasts are represented in the form of scenarios. It is assumed that from times \( t = 1 \) to \( t_1 \) demand \( d_{t}^k \) is known. After this demand is uncertain and this uncertainty is described via scenarios \( k = 1 : K \) which describe both demand development uncertainty and seasonal variations and variations due to holidays/special events, etc.

- \( d_{t}^k \) - demand forecast at time \( t \) under scenario \( k \), where \( k = 0 \) for \( t = 1 : t_1 \) and \( k = 1 : K \) for \( t = t_1 + 1, ..., T \);
- \( v_{t}^k \) - demand which will not be satisfied at time \( t \) under scenario \( k \);
- \( c_v \) - opportunity cost of not meeting one unit of demand;
- \( p_k \) - probability/frequency of scenario \( k, k = 1 : K \);

**Network description.**

- \( a_t \) - amount of capacity necessary to satisfy one unit of demand at time \( t \), it can be variable due to introduction of new services;
- \( C \) - capacity of one cell;
- \( D \) - capacity of one TRX;
- \( \Delta_z \) - time necessary for preparation of cell agreement;
- \( c_z \) - cost of preparation of cell agreement;
- \( \Delta_y \) - time necessary for building of a cell;
- \( c_y \) - cost of building of a cell;
- \( \Delta_x \) - time necessary for activation of a cell;
- \( c_x \) - cost of activation of a cell;
- \( \Delta_w \) - time necessary for setting up of a new TRX;
- \( c_w \) - cost of setting up of a new TRX;
- \( \alpha_t \) - coefficient used to discount costs at time \( t \) to the present.

**Network dimension and dimensioning decisions.** During time horizon \( t = 1, ..., t_1 \) they are described by the following quantities which refer to the beginning of period \( t \) and which depend on demand scenario \( k \) where \( k = 0 \) for \( t = 1 : t_1 \) and \( k = 1 : K \) for \( t = t_1 + 1, ..., T \).

- \( X^k_t \) - amount of cells in the working condition;
- \( Y^k_t \) - amount of built, but not activated cells;
- \( Z^k_t \) - amount of cell agreements;
- \( W^k_t \) - amount of TRX;
- \( x^k_t \) - amount of new cells to be activated; for \( t < 1 \) this is a model input resulting from previous decisions;
- \( y^k_t \) - amount of new cells to be built;
- \( z^k_t \) - amount of new cell agreements to be prepared;
- \( w^k_t \) - amount of new TRX to be set up.
- \( Z_{\text{max}} \) - maximal amount of cell agreements to initiate at any given time period.

**Development of the network capacity during time.** The following equations describe the development of fully operational cells, built cells, cell agreements and TRX over time.

\[
X^k_t = X^k_{t-1} + x^k_{t-\Delta_z} \tag{53} \\
0 \leq x^k_t \leq Y^k_t \tag{54}
\]
\[ Y^k_t = Y^k_{t-1} + y^k_{t-\Delta_x} - x^k_{t-1} \]
\[ 0 \leq y^k_t \leq Z^k_t \]
\[ Z^k_t = Z^k_{t-1} + z^k_{t-\Delta_x} - y^k_{t-1} \]
\[ 0 \leq z^k_t \leq Z_{\max} \]
\[ W^k_t = W^k_{t-1} + w^k_{t-\Delta_w} \]
\[ w^k_t \geq 0 \]

where \( k = 0 \) for the time horizon \( t = 1 : t_1 \) when demand is assumed to be known and \( k = 1 : K \) for the time horizon \( t = t_1 + 1 : T \) when demand is uncertain. The following relations connect these time horizons.

\[ X^k_{t_1} = X^0_{t_1}, \ Y^k_{t_1} = Y^0_{t_1}, \ Z^k_{t_1} = Z^0_{t_1}, \ W^k_{t_1} = W^0_{t_1}, \ k = 1 : K \]
\[ x^k_{t-\Delta_x} = x^0_{t-\Delta_x}, \ t_1 \leq t \leq t_1 + \Delta_x, \ k = 1 : K \]
\[ y^k_{t-\Delta_y} = y^0_{t-\Delta_y}, \ t_1 \leq t \leq t_1 + \Delta_y, \ k = 1 : K \]
\[ z^k_{t-\Delta_z} = z^0_{t-\Delta_z}, \ t_1 \leq t \leq t_1 + \Delta_z, \ k = 1 : K \]
\[ w^k_{t-\Delta_w} = w^0_{t-\Delta_w}, \ t_1 \leq t \leq t_1 + \Delta_w, \ k = 1 : K \]

**Relation between capacity and satisfied demand:**

\[ DW^k_t + a_t v^k_t \geq a_t d^k_t, \ v^k_t \geq 0 \]
\[ D (W^k_t + w^k_t) \leq C X^k_t. \]

**Costs.** Two types of costs are considered: the costs of network expansion and opportunity costs of not meeting demand. Additional costs could be considered, e.g. maintenance of cells and cell agreements.

**Capacity expansion decision.** This decision is taken at the beginning of time period \( t = 1 \) from the point of view of minimization of total costs which include costs of network expansion and opportunity costs of not meeting demand during time horizon \( t = 1 : t_1 \) plus costs of further network expansion contingent on demand scenarios during horizon \( t = t_1 + 1 : T \) averaged over scenarios and discounted to time \( t = 1 \). More formally:

Find \( x^0_t, y^0_t, z^0_t, w^0_t, X^0_t, Y^0_t, Z^0_t, W^0_t, v^0_t, t = 1 : t_1 \) from the solution of

\[
\min_{x^0_t, y^0_t, z^0_t, w^0_t, X^0_t, Y^0_t, Z^0_t, W^0_t, v^0_t} \sum_{t=1}^{t_1} \alpha_t \left( c_x x^0_t + c_y y^0_t + c_z z^0_t + c_w w^0_t \right) + \sum_{t=1}^{t_1} \alpha_t c_v v^0_t +
\sum_{k=1}^{K} p_k Q^k(x^0_t, y^0_t, z^0_t, w^0_t, X^0_t, Y^0_t, Z^0_t, W^0_t) \]

subject to (53)-(60),(66),(67) where \( k = 0 \). The arguments of the function \( Q^k(\cdot) \) represent the vectors with components indexed by time, for example \( x^0 = (x^0_1, x^0_2, ..., x^0_{t_1}) \). The function \( Q^k(\cdot) \)
represents the optimal costs of further network expansion during time horizon \( t = t_1 + 1 : T \) for a given demand scenario \( k \). In stochastic programming terminology this is a recourse problem

\[
Q^k(x^0, y^0, z^0, w^0, X^0, Y^0, Z^0, W^0) = \min_{x^k, y^k, z^k, w^k, X^k, Y^k, Z^k, W^k} \sum_{t=t_1+1}^{T} \alpha_t \left( c_x x^k_t + c_y y^k_t + c_z z^k_t + c_w w^k_t \right) + \sum_{t=t_1+1}^{T} \alpha_t c_v v^k_t
\]

subject to constraints (53)-(67). Problems (68)-(69) can be transformed into their deterministic equivalent and solved by linear programming software. We utilized the capabilities present in Excel spreadsheet for its solution.

The purpose of the model described above is to provide advice about aggregated decision of the investment into network expansion. It includes aggregated network characteristics and capacity constraints (67) which describe the capacity of the whole network. The inherently integer variables, like the number of cells or the number of TRX were substituted by their continuous approximations. In a more detailed model constraints (67) should be considered for a group of similar cells, or even for a given cell.

4 Enterprise level

The solutions of the network design problems considered in the previous section depend on a number of external parameters which were considered to be fixed. Some of these parameters derive from decisions which can be taken only on the strategic level through consideration of the whole environment in which the enterprise operates. Examples of such parameters are pricing of services, the total amount of resources allocated for investment and other quantities which characterise the strategy of the enterprise. The importance of this enterprise decision level has grown considerably during the recent years due to the increase in complexity of the telecommunications environment due to the convergence process with computer industry and content provision, deregulation and technological development. It is characterized by a growing uncertainty whose main sources are unpredictable market response to the new technologies and services, decreasing life cycles of products, actions of competition. Stochastic programming models represent a natural methodology for decision support on this level. However, they should be enhanced by selected ideas from the game theory in order to be capable of reflecting uncertainty which stems from actions of other decision makers.

4.1 Network operators and virtual providers: service pricing

The model developed refers to the situation represented on Figure 3. The environment consists of the Network Operator (NO), Virtual Service Provider (VNO) and a population of customers. Both NO and VNO provide a service to customers which can decide to subscribe to this service, change provider or discontinue to use the service altogether. Service providers decide the price for their service. The Network Operator possesses the network which is necessary for the service provision. The Virtual Network Operator does not have a network and in order to provide the service he has to lease the necessary network capacity from the Network Operator. Therefore the VNO has to decide how much capacity to lease and the NO decides the price to charge for the network capacity. Regulatory bodies may impose bounds on the leasing prices. Besides, the VNO
may provide additional value to a service which may lead to a market expansion which, in its turn, may make the existing network capacity inadequate. Therefore the network operator may face the necessity to invest in the network expansion.

The important question is what should be the bounds imposed by regulatory authorities on the price for leasing of the network capacity. From one side they should not be too high in order to permit the VNO to compete with the NO on the service provision. From the other side they should not be too low in order to permit the Network Operator to recover investment expenditures. We develop a stochastic optimization model which helps to answer this question. We take the point of view of the Network Operator which pursues the objective of maximizing his profit by making decisions about service prices, leasing prices and investment in the network. His decision model should contain the submodels which describe reactions of customers and the VNO to his policies.

Consider two decision periods $t = 1, 2$. Decisions are taken and implemented at the beginning of each period. During the rest of periods the market and competition reactions are observed which influence revenues and profits.

Decisions of NO and VNO. They are denoted by vectors $y_t$ and $z_t$ for NO and VNO respectively, $t = 1, 2$ where $y_1 = (y_{11}, y_{12}, y_{13})$, $y_2 = (y_{21}, ..., y_{25})$, $z_t = (z_{t1}, z_{t2})$. Components of these vectors have the following meaning.

$y_{11}, z_{t1}$ - prices charged for one unit of service by NO and VNO respectively during Period $t$.

$y_{12}$ - the price charged by NO for one unit of leased capacity.

$y_{13}$ - the maximal amount of capacity that NO is willing to lease.

$y_{24}$ - the amount of new capacity added to the network at the beginning of Period 2.

$y_{25}$ - the binary variable which equals 1 if decision is taken to expand the network capacity and zero otherwise.

$z_{t2}$ - the amount of capacity that VNO decides to lease.

Profit model of Network Operator for Period 1. The Network Operator takes the decision about prices at the beginning of the Period 1 without knowing precisely the reaction of customers and competition. His objective is to maximize the average profit taking into account the possibilities of profit and investment during the second period. The profit of NO is the difference between average revenue and costs during Period 1 with added discounted averaged profit from the Period 2. In order to describe it formally we need the following notations where we shall associate NO with

Figure 3: Competition between Network Operator and Virtual Network Operator
index $i = 1$ and VNO with index $i = 2$.

$\omega_i$ - vector of parameters which are uncertain for NO at the beginning of Period $t$. These parameters describe the market and competition reaction and will be defined more precisely when the market and competition models will be defined. It is enough that they have a known probabilistic description either in the form of continuous probability distributions or in the form of scenarios with given probabilities.

d$_{i,t}$ - demand for service provided by operator $i$ during Period $t$ measured by the network capacity required for its satisfaction.
a - network capacity owned by NO and available for service provision at the beginning of Period 1.
e$_{i,t}$ - cost of serving a unit of demand for operator $i$ during Period $t$.
g$_{i,t}$ - opportunity cost of not meeting one unit of demand for operator $i$ during Period $t$.
h$_{i,t}$ - cost of maintenance of one unit of capacity for operator $i$ during Period $t$; it is owned capacity for NO and leased capacity for VNO.
b - fixed costs associated with network expansion.
b$_{v}$ - variable costs per unit of capacity associated with network expansion.
$\alpha$ - coefficient for discounting the second period profit to the beginning of Period 1.

The profit of the NO is

$$F_{11}(y_1, z_1, d_{11}) = (y_{11} - e_{11}) \min \{d_{11}, a - z_{12}\} + y_{12} z_{12} - g_{11} \max \{0, d_{11} - a + z_{12}\} - ah_{11} + \alpha E_{\omega_1, \omega_2} Q(y_1, z_1, \omega_1, \omega_2)$$

(70)

where $Q(y_1, z_1, \omega_1, \omega_2)$ is the profit of the NO during Period 2, it depends on the decisions of both NO and VNO during Period 1 and on uncertain parameters $\omega_1, \omega_2$. The profit of the NO during Period 1 is described by the first four terms in (70). The first term represents the profit due to provision of service to customers. The second term represents the profit due to leasing of capacity to the VNO. The third term represents the opportunity costs of not meeting the demand for service provision while the fourth term does not depend on decision variables and represents the variable network maintenance cost.

An important feature of this profit model which distinguishes it from the models of the previous sections is the presence of two unknown components: decisions $z_1$ of the VNO and service demand $d_{11}$. To make this model useful for decision making the NO has to predict both these quantities. Such predictions are the scope of the market model and competition model which the NO should have. The market model provides the prediction of demand $d_{11} = d_{11}(y_1, z_1, \omega_1)$ for the service provided by both operators as function of their pricing decisions and uncertain parameters. The competition model provides the prediction of decisions $z_1 = z_1(y_1, \omega_1)$ of the VNO as function of decisions of the NO and uncertain parameters. Demand predictions $d_{12} = d_{12}(y_1, z_1, \omega_1)$ for the service provided by the VNO are used for making this prediction. Substituting these predictions into (70) we obtain the profit expression which depends only on the decisions of the NO and on uncertain parameters $\omega_1$ with known probabilistic description. The policy recommendation $y_1$ for the NO can be obtained by finding the values of $y_1$ which yield the highest mean profit. This leads to the following stochastic optimization problem:

$$\max_{y_1} E_{\omega_1} F_{11}(y_1, z_1(y_1, \omega_1), d_{11}(y_1, z_1(y_1, \omega_1), \omega_1))$$

(71)

$$y_{11} \leq y_{11} \leq y_{11}^+$$

(72)
\[ y_{12}^\leq y_{12} \leq y_{12}^+ \] (73)
\[ 0 \leq y_{13} \leq a \] (74)

where \( y_{11}^\leq, y_{11}^+, y_{12}, y_{12}^+ \) are price constraints imposed by regulatory and other considerations. We now give examples of the market model and the competition model.

**Market model.** The simplest such model ignores dependence of demand on prices and considers a finite number of demand scenarios with given probabilities. This is, however, an inadequate description of demand. A step towards more realistic demand representation is in the linear autoregressive model:

\[
d_{t1} = \max \left\{ 0, d_{t-1,1} + d_{11}^0 + r_{t1}(y_{t-1,1} - y_{t1}) + q_{t1}(z_{t1} - y_{t1}) \right\} \tag{75}
\]
\[
d_{t2} = \max \left\{ 0, d_{t-1,2} + d_{12}^0 + r_{t2}(z_{t-1,1} - z_{t1}) + q_{t1}(y_{t1} - z_{t1}) \right\} \tag{76}
\]

where \( t = 1, 2 \). Here \( d_{01} \) is demand for the service of operator \( i \) prior to the beginning of Period 1, it is assumed to be known; \( y_{01} \) and \( z_{01} \) are some initial reference service prices for both operators, \( d_{11}^0 \) is the component of the demand change for operator \( i \) during Period \( t \) which is not related to the price changes; \( r_{t1} \) is an additional demand obtained/lost due to a unit change of the price of operator \( i \) and \( q_{t1} \) is the flow of demand between the operators caused by a unit price difference between them. Parameters \( d_{11}^0, r_{t1}, q_{t1} \) are not known with certainty to the NO and constitute part of the vector \( \omega_t \) of unknown parameters. We have used the linear model due to its simplicity, but nonlinear models are also possible.

**Competition model.** This model summarizes the knowledge which the NO has about the objectives of the VNO. In the simplest case it is assumed that the VNO at time period \( t \) wants to maximize his expected current profit \( F_{t2}(y_t, z_t, d_{t2}) \) which has the structure similar to the profit of NO:

\[
F_{t2}(y_t, z_t, d_{t2}) = (z_{t1} - e_{t2}) \min \{ d_{t2}, z_{t2} \} - (y_{t2} + h_{t2}) z_{t2} - g_{t2} \max \{ 0, d_{t2} - z_{t2} \} \tag{77}
\]

where the first term represents the profit derived from the service provision, the second one reflects expenditures due to the network leasing and the third one reflects opportunity costs for not meeting demand. Parameters \( e_{t2}, h_{t2} \) and \( g_{t2} \) are uncertain for the NO and represent another part of components of the vector of uncertain parameters \( \omega_t \). The dependence of demand \( d_{t2} \) on prices of both operators is obtained from the market model (75)-(76) and substituted into (77). After this the prediction \( z_t(y_t, \omega_t) \) of response of the VNO to decisions of the NO is obtained as solution of

\[
\min_{z_t} F_{t2}(y_t, z_t, d_{t2}(y_t, z_t, \omega_t)) \tag{78}
\]

\[
z_{t1}^\leq \leq z_{t1} \leq z_{t1}^+ \tag{79}
\]
\[
0 \leq z_{t2} \leq y_{t3} \tag{80}
\]

where \( z_{t1}^-, z_{t1}^+ \) are price bounds imposed by regulation and other considerations.

**Profit model of Network Operator for Period 2.** This model reflects the future profits \( Q(y_1, z_1, \omega_1, \omega_2) \) in the total profit of the NO. This will make the first period decision more forward looking and capable to facilitate an eventual adaptation to changing market circumstances by investment in the network expansion. The profit of the NO during Period 2 is

\[
F_{21}(y_2, z_2, d_{21}, \omega_2) = (y_{21} - e_{21}) \min \{ d_{21}, a - z_{22} + y_{24} \} + y_{22} z_{22} - \\
g_{21} \max \{ 0, d_{21} - a + z_{22} - y_{24} \} - h_{21} (a + y_{24}) - b_w y_{24} - b_f y_{25} \tag{81}
\]

28
where the network expansion occurs at the beginning of the Period 2 and the total network capacity available for service provision during Period 2 is \( a + y_{24} \). This expression for the profit is similar to (70) and contains two new last terms which reflect variable and fixed costs related to the expansion of network. Parameters \( e_{21}, g_{21}, h_{21}, b_v, b_f \) can be uncertain for the NO at the beginning of Period 1 and constitute additional components of the vector \( \omega_2 \), others being the components similar to those of \( \omega_1 \). Prediction \( d_{21} = d_{21}(y_2, z_2, \omega_2) \) of demand is obtained from the demand model (75)-(76) and prediction \( z_2(y_2, \omega_2) \) of competition response is obtained from the competition model (78)-(80). These predictions depend also on \( y_1, z_1, \omega_1 \) through autoregressive relations (75)-(76), but we omitted this dependence to simplify notations. The value of the future profit is obtained by solving

\[
Q(y_1, z_1, \omega_1, \omega_2) = \min_{y_2} F_{21}(y_2, z_2(y_2, \omega_2), d_{21}(y_2, z_2, \omega_2), \omega_2)
\]

where (83)-(85) are similar to constraints (72)-(74). Constraint (85) reflects the assumption that the NO can not offer to VNO less capacity during Period 2 than the amount offered by him during Period 1. Constraint (86) will force the amount of the network extension to zero if the decision not to expand the network was taken. Otherwise it will limit the network expansion to the maximal admissible level \( M \).

The problem (71)-(74) is an extension of the classical stochastic programming problem with recourse to the case when the part of the uncertainty is due to the actions of other decision makers. This adds a new level of complexity in the form of prediction problem (78)-(80). The problem (82)-(86) is an extension of the classical recourse problem with the new feature added by prediction problem (78)-(80). This additional complexity makes the problem far more challenging than traditional linear stochastic problems with recourse. The linearity structure is never present here. However, some general approaches still can be used, notably the transformation of the problem into its deterministic equivalent by representing the uncertain parameters \( \omega_1, \omega_2 \) through a finite number of scenarios. Nonlinear programming software can be used for solution of such deterministic equivalent. Another promising approach is the stochastic quasigradient methods (Gaivoronski, 1988).

We finish this section by presenting on Figure 4 a typical example of a computation of the profit function \( F_1(y_1) = E_{\omega_1} F_{11}(y_1, z_1(y_1, \omega_1), d_1(y_1, z_1(y_1, \omega_1), \omega_1)) \) of the Network Operator performed by Matlab 6.1 with Optimization Toolbox.

On this table the vertical axis marked by \( F_1(y) \) shows the values of the profit function of the NO computed according to (70). The two horizontal axes marked by \( y_1 \) and \( y_2 \) show the values of the service price and leasing price respectively of the NO. The figure shows the complex nature of the profit function which exhibits different patterns in different regions of the price space. This space can be divided into four regions. In the first region both service price and leasing price are moderate which results in the pattern where both operators have positive share in service market. In the second region the service price is high while the leasing price is moderate. In this region the NO has no customers and gets all his profit from leasing the capacity to the VNO which monopolizes the service market. The opposite picture can be observed in the third region where the leasing price
is high while the service price is moderate. In this region the NO becomes a monopolist in service provision and the VNO is squeezed out from the market. Finally, in the fourth region where both prices are high the market does not take off at all because the high service price discourages the customers from subscribing to the service of the NO, while the high leasing price prevents the VNO from offering the service at an attractive price. The good news which can be derived from this and similar examples Audestad et al. (2002) is that despite its complex nature the profit function has a distinctive structure and within each of the regions its behavior is close to concave. This circumstance can be exploited in numerical methods.

5 Summary

In this paper we gave a survey of different applications of stochastic optimization to telecommunications. Some presented models are new, while others are close to those found in the literature. Stochastic programming is a methodology of choice for support of complex network design decisions in the presence of uncertainty. In telecommunications uncertainty is present on all levels of network design, starting from the level of technology, through the level of network design to the enterprise level where the top level strategic decisions are taken. Moreover, due to the current changes in the industry environment the adequate treatment of uncertainty is becoming paramount for taking competitive design and investment decisions. As different examples in this paper show, accumulated modeling and computational experience in solving stochastic optimization problems represent a solid background for deployment of stochastic programming applications in telecommunications. At the same time more work is needed in development of methodology, in particular in nonlinear and mixed integer stochastic programming models.
Acknowledgement

Thanks are due to Dr. Mario Bonatti of Italtel and Dr. Jan-Arild Audestad of Telenor for useful discussions which helped to shape some of the ideas in this paper.

References


